

BAB II

LANDASAN TEORI

2.1 *Stroke*

Beberapa hal yang perlu diperhatikan untuk memahami *Stroke* diantaranya:

2.1.1 Definisi *Stroke*

Stroke adalah penyakit yang menyerang jaringan otak yang disebabkan berkurangnya aliran darah dan oksigen ke dalam otak. Berkurangnya aliran darah dan oksigen ini disebabkan karena adanya sumbatan, penyempitan, atau pecahnya pembuluh darah di dalam otak tersebut (Okthavia, 2014).

Stroke diartikan oleh orang awam dengan istilah penyakit lumpuh, padahal *Stroke* tidak selalu disertai dengan kelumpuhan. *Stroke* juga disebut serangan otak. *Stroke* adalah suatu serangan pada otak akibat gangguan pembuluh darah dalam mensuplai darah yang membawa oksigen dan glukosa untuk metabolisme sel-sel otak agar dapat tetap melaksanakan fungsinya. *Stroke* adalah suatu kondisi yang ditandai dengan serangan otak akibat pukulan telak yang terjadi secara mendadak. Dalam bahasa medis *Stroke* disebut CVA (*Cerebro Vascular Accident*). Merujuk pada istilah medis, *Stroke* didefinisikan sebagai gangguan saraf permanen akibat terganggunya peredaran darah ke otak, yang terjadi sekitar 24 jam atau lebih. Sindrom klinis ini terjadi secara mendadak serta bersifat progresif sehingga menimbulkan kerusakan otak secara akut dengan tanda klinis yang terjadi secara lokal dan atau global (Novida & Santi, 2014).

2.1.2 Jenis-jenis *Stroke*

Penyakit *Stroke* dapat dibedakan berdasarkan jenisnya menjadi (Caplan, 2006):

1. *Stroke* Iskemik (*Ischemic Stroke*), didefinisikan sebagai *Stroke* akibat aliran darah otak yang berkurang sehingga jaringan otak mengalami defisit nutrisi berupa oksigen dan glukosa untuk bermetabolisme. *Stroke* Iskemik mengakibatkan suatu cedera yang disebut dengan infark (*infarction*).

Hak Cipta Dilindungi Undang-Undang

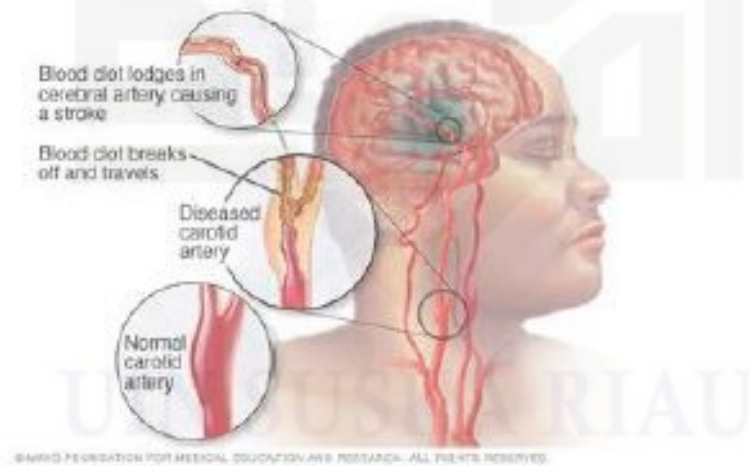
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Stroke Iskemik ini dapat sembuh dalam beberapa minggu dan sering disebut dengan *Stroke* Ringan. *Stroke* iskemik dibagi menjadi 3 kelompok berdasarkan patofisiologinya menjadi :

- a. *Stroke* Trombotik (*Thrombosis Stroke*), merupakan *Stroke* akibat sumbatan pada salah satu atau lebih pembuluh darah yang dikarenakan adanya pembentukan lapisan dalam pada pembuluh darah otak sehingga lumen pembuluh menyempit dan aliran darah menjadi terhalang.
- b. *Stroke* Embotik (*Embolism Stroke*), merupakan *Stroke* yang disebabkan oleh terbentuknya material embolus yaitu benda asing seperti *thrombus* atau udara pada pembuluh darah otak sehingga menyumbat aliran darah otak.
- c. Hipoperfusi Sistemik (*Systemic Hypoperfusion*), merupakan *Stroke* akibat menurunnya kerja organ sistemik seperti contohnya kegagalan pompa pada jantung dan hipotensi akibat kehilangan sejumlah darah sehingga aliran darah menuju otak menjadi berkurang.



Gambar 2. 1 Ilustrasi *Stroke* Iskemik Trombotik
(Sumber : Mayo Medical Education, 2006)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

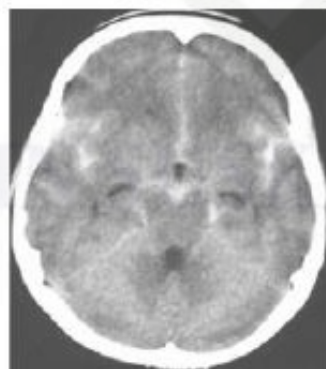
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 2. 2 Citra CT Scan Stroke Iskemik
(Sumber : Fisika, 2013)

Stroke Hemoragik (Hemorrhage Stroke), didefinisikan sebagai *Stroke* akibat adanya kebocoran darah ke ruang *ekstravaskular* pada *kranium* atau pendarahan. *Stroke* hemoragik ini sembuh dengan meninggalkan cacat, tidak bisa sembuh total, bahkan dalam beberapa bulan atau tahun dan dapat menyebabkan kematiandan sering disebut dengan *Stroke Berat*. *Stroke* hemoragik dikelompokkan menjadi 2 jenis yaitu :

- Pendarahan Subarakhnoid (*Subarachnoid Hemorrhage*), merupakan pendarahan pada pembuluh darah otak sehingga darah mengalami kebocoran menuju permukaan otak dan masuk ke ruang Subarakhnoid. Darah yang bocor akan meningkatkan tekanan darah dan tekanan intrakradinal.



Gambar 2. 3 Citra CT Scan Stroke Hemoragik Subarakhnoid
(Sumber : Fisika, 2013)

- Pendarahan Intracerebral (*Intracerebral Hemorrhage*), merupakan pendarahan pada pembuluh darah otak yang langsung masuk pada

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

jaringan otak. Derajat keparahan dari pendarahan ini tergantung lokasi pendarahan, kecepatan kebocoran, volume dan tekanan darah akibat pendarahan yang menyebabkan edema otak.



Gambar 2. 4 Citra CT Scan Stroke Hemoragik Intraserebra
(Sumber : Fisika, 2013)

Sedangkan berdasarkan gejalanya *Stroke* dibagi menjadi dua, yaitu :

1. *Stroke Berat*, *Stroke* ini sembuh dengan meninggalkan cacat, tidak bisa sembuh total, bahkan dalam beberapa bulan atau tahun dan dapat menyebabkan kematian. Yang termasuk stroke berat adalah *Stroke Iskemik*.
2. *Stroke Ringan*, *Stroke* ini dapat sembuh dalam beberapa minggu.

2.1.3 Faktor-faktor *Stroke*

Terdapat beberapa faktor resiko pada penyakit *Stroke* menurut (Haris & Kurnia, 2015), yaitu:

1. Usia

Semakin tua usia, maka resiko untuk terjadinya *Stroke* juga semakin meningkat. Usia tua juga dihubungkan sebagai faktor resiko *Stroke* dari faktor resiko *Stroke* yang lainnya yaitu berupa Hipertensi, Diabetes Mellitus, dan Kelainan Jantung.

2. Tekanan Darah

Tekanan Darah dapat menyebabkan pecahnya pembuluh darah dari otak sehingga dapat menyebabkan pendarahan intraserebral, intraventrikular, maupun subaraknoid.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3. Kadar Gula Darah

Diabetes Mellitus meningkatkan resiko *Stroke* menjadi 1-3 kali lipat bila dibandingkan dengan bukan penderita Diabetes Mellitus. Walaupun Diabetes Mellitus bukan merupakan faktor independen penyebab *Stroke*, namun pengendalian kadar gula dapat mengurangi komplikasi pada pembuluh darah yang nantinya akan berperan dalam kejadian *Stroke*.

4. Kadar Kolesterol

Sama halnya dengan Diabetes Mellitus, hiperlipidemia juga bukan merupakan independen penyebab *Stroke* meskipun dalam beberapa penelitian menyebutkan bahwa dengan menurunnya kadar kolesterol darah maka resiko untuk terkena penyakit *Stroke* juga akan menurun.

5. Riwayat penyakit Jantung

Dalam penyakit jantung terdapat pembuluh antara lain atrial fibrilasi dan stenosis arteri karotis. Atrial Fibrilasi meningkatkan resiko terjadinya *Stroke* dua kali lipat pada laki-laki dan tiga kali lipat pada perempuan. Penyakit jantung menjadi penyebab terjadinya langsung serangan *Stroke*. Pada *Stroke* yang disebabkan oleh emboli, faktor yang paling berperan adalah infark miokard, aritmia, dan atrial fibrilasi.

6. Indeks Massa Tubuh

Obesitas dapat meningkatkan resiko penyakit *Stroke* 2 kali lipat bila dibandingkan dengan bukan penderita *Obesitas*. Diketahui pula efek dari *Obesitas* adalah mempercepat aterosklerosis pada remaja dan dewasa muda. Oleh karena itu, penurunan berat badan dapat mengurangi resiko terserang *Stroke*.

7. Perilaku Merokok

Merokok dapat menyebabkan peningkatan koagulabilitas, viskositas darah, meningkatkan kadar-kadar fibrinogen, memacu agregasi trombosit sehingga meningkatkan tekanan darah, dan dapat meningkatkan hematokrot. Orang-orang yang memiliki kebiasaan merokok cenderung lebih beresiko untuk terkena penyakit jantung dan *Stroke* dibandingkan

orang yang tidak merokok. Merokok meningkatkan resiko *Stroke* sebesar 1,5 kali sesudah faktor resiko *Stroke* yang lain.

8. Aktivitas Fisik

Meningkatkan aktivitas fisik akan menurunkan resiko *Stroke*. Rendahnya aktivitas fisik akan menjadi *predictor* faktor resiko *Stroke* Iskemik. Aktivitas fisik berperan sebagai efek protektif pada *Stroke* dan penyakit Kardiovaskuler lain. Olahraga aerobik teratur berpengaruh positif terhadap pencegahan *Stroke* melalui mekanisme pengendalian berat badan, tekanan darah, dan kadar kolesterol dalam darah.

2.1.4 Gejala *Stroke*

Data gejala atau anamnesis yang digunakan oleh ahli saraf dalam mendiagnosis jenis *Stroke* yang diderita adalah :

1. Otot wajah lemah, gejala ini ditunjukkan oleh wajah yang menjadi tidak simetris, bagian mata dan mulut turun yang menunjukkan adanya lesi pada saraf.
2. Gangguan bicara, gejala ini di tunjukan adanya gangguan bicara (pelo) dan mengerti pembicaraan.
3. Lemah pada satu sisi tubuh. Tangan, kaki dan wajah tiba-tiba lemah.
4. Gangguan sensorik, yaitu berkurangnya sensibilitas pada permukaan kulit tubuh pasien. Gejala ini berupa :
 - *Hemihipesthesia*, berupa berkurangnya sensitivitas pada satu sisi tubuh.
 - *Hemianesthesia*, berupa mati rasa pada satu sisi tubuh.
 - *Hemiparesthesia*, berupa rasa kesemutan pada satu sisi tubuh.
5. Nyeri kepala, gejala ini bisa diikuti dengan mual, muntah dan kejang.

2.2 KDD (*Knowledge Discovery In Database*)

Knowledge Discovery In Database (KDD) merupakan salah satu metode untuk memperoleh suatu pengetahuan dari *database* yang ada. Hasil pengetahuan yang diperoleh dari data tersebut dapat digunakan sebagai basis pengetahuan (*knowledge base*) untuk keperluan mengambil keputusan (Mardi, 2015). KDD merupakan keseluruhan proses konversi data mentah menjadi pengetahuan yang

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

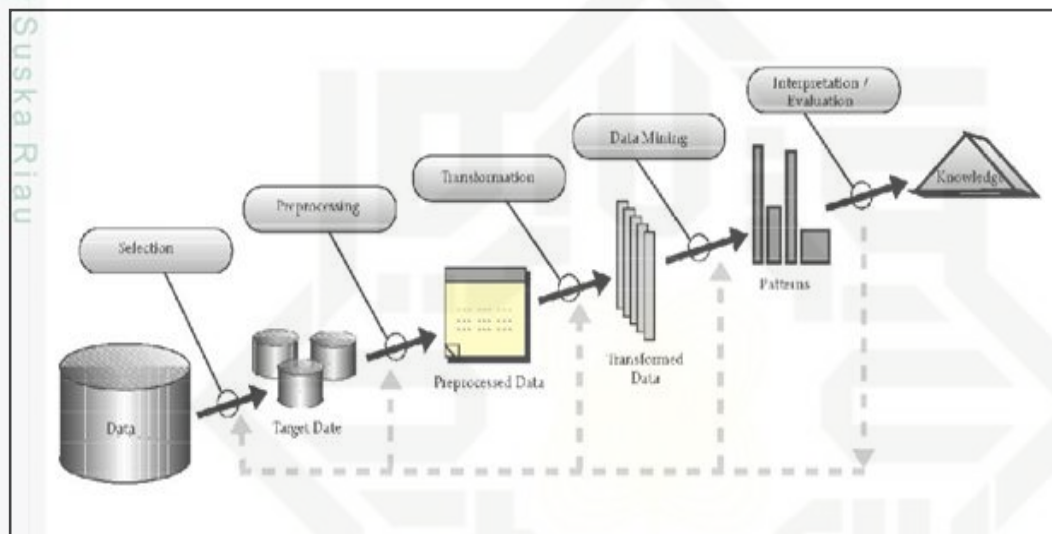
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

bermanfaat yang terdiri dari serangkaian tahap. KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan pola atau hubungan dalam set data ukuran besar (Muslehatin et al., 2017).

Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Mardi, 2015):



Gambar 2. 5 Tahapan KDD (Sumber Fayyad dkk,1996)

Tahapan-tahapan keseluruhan dari proses KDD adalah:

1. *Data Selection*

Data seleksi adalah pemilihan seleksi data dari sekumpulan data operasional perlu dilakukan sebelum tahapan penggalian informasi dalam KDD dimulai. Seleksi digunakan untuk menentukan variabel apa saja yang akan diambil agar tidak terjadi kesamaan dan perulangan yang tidak diperlukan dalam pengolahan teknik *data mining*.

2. *Preprocessing* Pada *Preprocessing* terdapat dua tahap, yaitu sebagai berikut:

a. *Data Cleaning*

Untuk menghilangkan data yang tidak diperlukan seperti menangani nilai yang hilang (*missing value*), data *noise* dan menangani data yang tidak konsisten dan relevan.

1. Menangani *missing value* dengan cara :

- Abaikan *record*.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Masukkan nilai yang hilang secara manual.
 - Gunakan konstanta umum untuk mengganti nilai yang hilang.
 - Gunakan rata-rata nilai atribut untuk mengganti nilai yang hilang.
 - Gunakan rata-rata nilai atribut dari semua sampel yang berada pada kelas yang sama.
 - Gunakan nilai yang paling mungkin diganti untuk nilai yang hilang.
2. Penanganan *Noisy* data dengan cara :
 - *Smoothing* (penghalusan data), berkonsultasi dengan tetangga secara lokal.
 - Mendekati *outlier* atau diluar nilai wajar.
 3. *Inconsisten* data dilakukan dengan manual menggunakan referensi-referensi eksternal.

b. *Data Integration*

Data Integration dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti nama, jenis produk dll.

3. Transformasi

Perubahan data menjadi format ekstensi yang sesuai untuk pengolahan dalam *data mining*. Beberapa metode *data mining* memerlukan format yang khusus sebelum bisa diproses pada tahapan *data mining*. Misalnya sebagai metode standar analisa *Clustering* dan asosiasi hanya bisa menerima data informasi kategorikal.

4. *Data mining*

Saat proses *data mining* teknik utama metode diterapkan agar menemukan pengetahuan berharga, data yang terkumpul sesuai prosedur harus diterapkan pada proses *mining* setelah data melalui transformasi. Ada berbagai macam teknik *data mining* seperti *estimation*, *prediction*, *classification* dan *association*.

5. *Evaluation / Interpretation*

Tahapan ini yaitu mengidentifikasi pola-pola menarik kedalam *knowledge based* yang diidentifikasi. Dalam tahap ini hasil dari teknik *data mining*

berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah kajian yang ada sudah memenuhi target yang diinginkan. Jika ternyata hasil yang diperoleh tidak sesuai kajian ada beberapa alternatif dengan mencoba metode *data mining* lain agar lebih sesuai.

6. Knowledge

Pola-pola yang dihasilkan akan dipresentasikan kepada pengguna. Pada tahapan ini pengetahuan baru yang dihasilkan bisa dipahami semua orang yang akan dijadikan acuan pengambilan keputusan.

2.3 Data Mining

Data mining adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. *Data mining* merupakan serangkaian proses untuk mengenali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual (Mardi, 2015).

Menurut *Garthener Group*, *data mining* adalah proses menemukan hubungan baru yang mempunyai arti, pola dan kebiasaan dengan memilah-milah sebagaimana besar data yang disimpan dalam media penyimpanan dengan menggunakan teknologi pengenalan pola seperti teknik statistik dan matematika (Arifianto & Sarosa, 2014). *Data mining* menurut David Hand, Heikki Mannila, dan Pedhraic Smyth dari MIT adalah analisa terhadap data (biasanya data yang berukuran besar) untuk menemukan hubungan yang jelas serta menyimpulkan yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut.

2.3.1 Karakteristik Data Mining

Data mining dibagi menjadi beberapa metode berdasarkan tugas yang dapat dilakukan yaitu :

1. Asosiasi (*Association*)

Digunakan untuk mengenali kejadian-kejadian khusus atau proses dimana hubungan asosiasi muncul pada setiap kejadian.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.3 Pengklusteran (*Clustering*)

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau mempertahankan dan membentuk kelas objek-objek yang memiliki kemiripan.

3. Prediksi (*Prediction*)

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai hasil akan ada dimasa mendatang.

4. Estimasi (*Estimation*)

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik dari pada kearah kategori.

5. Klasifikasi (*Classification*)

Klasifikasi adalah fungsi pelajaran yang memetakan (mengklasifikasi) sebuah unsur (*item*) data ke dalam salah satu dari beberapa kelas yang sudah didefinisikan.

2.4 Naïve Bayes

Naïve Bayes merupakan salah satu metode klasifikasi dalam *data mining* berdasarkan probabilitas dan Teorema *Bayes*. Adapun asumsi yang harus dipenuhi yaitu independensi dalam variabel bebas (Septiani, 2017). Maksudnya adalah kemunculan kata tidak mempengaruhi kemunculan kata lainnya, atau sebaliknya. Algoritma *Naïve Bayes* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasikan data uji pada kategori yang paling tepat (Ronen, 2007).

Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang ditemukan oleh ilmuan *Inggris Thomas Bayes*, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai *Thomas Bayes*. Teorema tersebut dikombinasikan dengan *Naïve* dimana diasumsikan kondisi antara atribut saling bebas. Klasifikasi *Naïve Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lain. Teorema ini mampu diaplikasikan dalam *database* dengan data yang besar yang memiliki tingkat akurasi dan

kecepatan yang tinggi. Berikut merupakan Teorema Bayesian dalam bentuk umum:

$$P(A|B) = \frac{(P(B|A) \cdot P(A))}{P(B)} \dots \dots \dots (2.1)$$

Keterangan :

B : Data dengan kelas yang belum diketahui.

A : Hipotesa data B merupakan suatu kelas spesifik.

P(A) : Probabilitas hipotesa A (*prior probability* / probabilitas awal).

P(B) : Probabilitas B.

P(B|A) : Probabilitas hipotesa B berdasarkan kondisi A.

P(A|B) : Probabilitas hipotesa A berdasarkan kondisi B (*Posterior probability* / probabilitas akhir).

Ide dasar dari aturan *Bayes* adalah hasil dari hipotesis atau peristiwa (A) dapat di prediksi berdasarkan bukti (B) yang di amati. Yang harus di perhatikan yang pertama, dalam sebuah probabilitas awal/prior (A) atau P(A) adalah sebuah probabilitas dari suatu hipotesis sebelum bukti dilakukan dan yang kedua, sebuah probabilitas akhir (A) atau P(A|B) adalah probabilitas dari suatu hipotesis setelah bukti dilakukan.

2.4.1 *Naïve Bayes* untuk klasifikasi

Kaitan pada *Naive Bayes* dan klasifikasi, korelasi hipotesis dan bukti klasifikasi bahwa hipotesis dalam teorema adalah label kelas yang menjadi tujuan dari pemetaan. Dalam *Naive Bayes Classifier* maksud independensi yang kuat pada atribut adalah sebuah atribut pada sebuah data tidak berkaitan dengan ada atau tidaknya atribut lain dalam data yang sama. Independensi (tidak ketergantungan), dengan kata lain *Naive Bayes Classifier* model yang digunakan adalah model atribut independensi (Prasetio, 2012).

Konsep probabilitas merupakan salah satu bentuk model statistik. Salah satu metode yang menggunakan konsep probabilitas adalah *Naive Bayesian Classification* (NBC). Pada metode ini, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain. Apabila diberikan k atribut

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

yang saling bebas (independence), nilai probabilitas dapat diberikan sebagai berikut. Klasifikasi *Naive Bayes Classifier* yang memiliki nilai atribut yang terpisah atau independen ini dapat dinyatakan dengan rumus sebagai berikut:

$$P(X_1, \dots, x_k | C) = P(X_1 | C) \times P(X_2 | C) \times \dots \times P(X_k | C) \dots \dots \dots (2.2)$$

Keterangan :

- $P(X_1, \dots, x_k | C)$ = Probabilitas kelas C menghasilkan intence X
- $P(X_1 | C)$ = Probabilitas kelas C menghasilkan nilai yang diamati untuk atribut ke-1 (X_1)
- $P(X_2 | C)$ = Probabilitas kelas C menghasilkan nilai yang diamati untuk atribut ke-1 (X_2)

Untuk menghitung nilai akhir kelas menggunakan rumus :

$$C_{MAP} = \operatorname{argmax}_{c \in C} P(X | C) \dots \dots \dots (2.3)$$

Keterangan :

- C_{MAP} = Hipotesa nilai tertinggi
- $\operatorname{argmax}_{c \in C}$ = Nilai rata-rata dari setiap kelas

2.4.2 Karakteristik *Naive Bayes Classifier*

Klasifikasi *Naive Bayes Classifier* bekerja berdasarkan probabilitas yang mengamati semua atribut data sebagai bukti dalam probabilitas. Hal ini dapat kita ketahui dari karakteristik *Naive Bayes Classifier* sebagai berikut:

- Metode *Naive Bayes Classifier* bekerja teguh terhadap data-data terisolasi yang biasanya merupakan data dengan karakteristik berbeda. *Naive Bayes Classifier* juga menangani nilai atribut yang salah dengan mengabaikan data latih selama proses pembangunan model dan prediksi.
- Kuat atau tangguh menghadapi atribut yang tidak relevan.
- Atribut yang mempunyai korelasi bisa mendegradasi kinerja klasifikasi *Naive Bayes Classifier* karena asumsi independensi atribut tersebut sudah tidak ada.

2.5 Evaluasi

Evaluasi merupakan kunci dalam pembuatan aplikasi yang berbasis *data mining*. *Confusion Matrix* adalah sebuah metode yang digunakan untuk menganalisa seberapa baik sebuah model klasifikasi yang dibangun dalam mengenali data dengan kelas-kelas yang berbeda. Dengan performa dari suatu model klasifikasi dapat diukur dengan tingkat akurasi. Akurasi yang dihasilkan dari sebuah klasifikasi yang memberikan hasil latihan dengan bentuk persentase dari kelompok data latihan yang telah diklasifikasikan dengan benar. Dengan perhitungan (Han & Kamber, 2012):

		Predicted class		
		yes	no	Total
Actual class	yes	TP	FN	P
	no	FP	TN	N
Total		P'	N'	P + N

Gambar 2. 6 Tabel Confusion Matrix (Han, Kamber dan Pei 2012)

Keterangan :

Actual class (yes) : kelas sebenarnya yang bernilai *yes*.

Actual class (no) : kelas sebenarnya yang bernilai *no*.

Predicted class (yes) : prediksi kelas yang bernilai *yes*.

Predicted class (no) : prediksi kelas yang bernilai *no*.

TP (*True Positive*) : jumlah data yang diklasifikasikan dengan benar (*Actual class (yes) = Predicted class (yes)*).

FN (*False Negative*) : jumlah data yang diklasifikasikan dengan salah (*Actual class (yes) = Predicted class (no)*).

FP (*False Positive*) : jumlah data yang diklasifikasikan dengan salah (*Actual class (no) = Predicted class (yes)*).

TN (*True Negative*) : jumlah data yang diklasifikasikan dengan benar (*Actual class (no) = Predicted class (no)*).

P : Jumlah dari TP dan FN

N : Jumlah dari FN dan TN

TP dan TN menunjukkan bahwa model klasifikasi melakukan klasifikasi dengan benar, FP dan FN menunjukkan bahwa model klasifikasi melakukan klasifikasi dengan salah. Sehingga untuk menghitung akurasi dari model yang dibangun dengan persamaan berikut :

$$Accuracy = \frac{TP+TN}{P+N} \times 100\% \dots\dots\dots(2.4)$$

2.6 Penelitian Terkait

Berikut ini merupakan beberapa penelitian terkait yang menjadi referensi pada tugas akhir ini. Adapun penelitian terkaitnya dapat dilihat pada tabel 2.1 berikut :

Tabel 2.1 Karya Penelitian Terkait

No	Judul Penelitian	Nama Peneliti	Asal	Tahun	Hasil
1	Klasifikasi Status Gizi Menggunakan <i>Naïve Bayes Classifier</i>	Kusuma Dewi	Universitas Islam Indonesia	2009	Naive Bayes Classifierian <i>Classification</i> (NBC) dapat digunakan sebagai salah satu metode untuk klasifikasi status gizi karena hasil pengujian menunjukkan total kinerja sebesar 0,932 atau 93,2%.
2	Tingkat Faktor Risiko Stroke dengan Pengetahuan Masyarakat terhadap Deteksi Dini Penyakit <i>Stroke</i> .	Haris Faisal dan Kurnia Rachmawati	Universitas Lampung Mangkurat	2015	Tingkat faktor risiko stroke rendah sebanyak 35%, tingkat faktor risiko stroke sedang sebanyak 37,5%, dan tingkat faktor risiko stroke tinggi sebanyak 27,5%. Tingka

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau State Islamic University of Sultan Syarif Kasim Riau

No	Judul Penelitian	Nama Peneliti	Asal	Tahun	Hasil
					pengetahuan responden terhadap deteksi dini penyakit stroke rendah sebanyak 15%, pengetahuan sedang sebanyak 45%, dan pengetahuan tinggi sebanyak 40%.
3	Penerapan Algoritma Naïve Bayes untuk Klasifikasi Data Nasabah Asuransi.	Bustami	Aceh	2014	Algoritma <i>Naive Bayes</i> di dukung oleh ilmu probabilistik dan ilmu statistika khususnya dalam penggunaan data petunjuk untuk mendukung keputusan pengklasifikasian. Pada algoritma <i>Naive Bayes</i> , semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut.
4	Perancangan Sistem Pakar Diagnosa Potensi Serangan <i>Stroke</i> menggunakan Metode Fuzzy	Budiarto	Medan	2015	Program sistem pakar ini digunakan untuk mendiagnosis seorang pasien terkena penyakit penyakit <i>Stroke</i> berdasarkan variabel data medis yang ditunjukkan pasien dan Sistem pakar sebagai alat

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

No	Judul Penelitian	Nama Peneliti	Asal	Tahun	Hasil
					bantu untuk membantu para Dokter baru dalam menentukan potensi penyakit <i>Stroke</i> dialami pasien, serta membantu masyarakat umum untuk mengetahui potensi terkena penyakit <i>Stroke</i> .